Statistics 10 Lecture 14 CHANCE ERRORS IN SAMPLING (Chapter 20.1-20.3)

1. Overview

Sample surveys involve chance error. The 2000 presidential election is a perfect example. Remember POPULATION and SAMPLE. No matter who the researcher is -- an individual, a company, a polling organization, the interest is in the PARAMETER, but given resource constraints, virtually everyone settles for a STATISTIC. The difference in the value between the PARAMETER and the STATISTIC is chance error if there is no bias.

2. A Basic Example (20.1)

Ask, what is the population and what was the sociologist expecting as the percentage of men in her sample of 100?

3. Sample Size and the Standard Error

If we increase the size of the sample (assuming it is representative) we get a better "fix" on the PARAMETER. In Figure 2 (p. 358) we draw 250 samples of size 400 and now the range is 39% men to a high of 54%. When she was drawing samples of size 100 the number of men in each sample ranged from a low of 34% to a high of 58%. The standard error got smaller as the sample size got larger. What does this mean for us?

4. The Standard Error for a Sample Percentage (20.2)

Remember the Standard Error for a Sum (17.2)? It was

Standard Error for a sum = $\sqrt{\begin{matrix}\textit{Sample}\\ \textit{Size}\end{matrix}}$ * Standard Deviation of "the box".

Example 1: for a sample of 100 from the sociologist's population, the standard error for the sum (or total) of men is: $\sqrt{100} * [(1-0)\sqrt{.46 * .54}\,] = 10 \times .5 = 5$

And the SE for the percentage of men = (SE for a sum / size of sample) x 100 = (5/100) * 100 = 5%

Example 2: for a sample of 400, the standard error is $\sqrt{400} * [(1-0)\sqrt{.46 * .54}\,] = 20 \times .5 = 10$

And the SE for a percentage = (SE for a sum / size of sample) x 100 = (10/400) = 2.5%

5 . Interpretation and the Normal Curve Again (20.3)

A third example: Problem 2, Exercise Set A : 25,000 students, 10,000 are older than 25.

Find the expected value for the total number of students older than age 25 in a sample of 400.

That will be = number of "draws" x average of a box or 160 = (400 x .4).

This is like a box with 10,000 1's and 15,000 0's . So Find the standard error of the number of students older than 25 in the sample = square root of sample size x SD of box = $\sqrt{400} * [(1-0)\sqrt{.40 * .60}\,] = 20 \times .5 = 10$.

Find the standard error of the percentage of students = (10 / 400) x 100 = 2.5%

So the percentage of students in the sample of 400 who are older than 25 will be around 40% give or take 2.5%. That is fine, but what does it mean?

One standard error in this example is 2.5%, +1 standard error would be 40% +2.5% or 42.5%, -1 standard error will be 37.5%. The chance that between 37.5% and 42.5% of any given sample of 400 students will be older than 25 is about 68%.

We can move from using the normal curve to figuring percentages to using the normal curve to make statements about chances.

Statistics 10 Lecture 14 CHANCE ERRORS IN SAMPLING (Chapter 20.1-20.3)

The chance that between 35% and 45% of any given sample of 400 students will be older than 25 is about 95%. And 99%? 32.5% to 47.5%

What's going on here? As was suggested in Chapter 17.3, 18, and 20.1, if we could sample infinitely, the sample percentages would bunch around the true value and have the appearance of a normal distribution.
We can borrow from that property and use the normal curve to make statements about chances of getting samples with particular characteristics. We can convert percentages (or any other kind of statistic) to Z scores. Remember this from Chapter 17.3?

Z = (observed sum - expected sum)

Standard Error of a Sum

In chapter 20.3, it now looks like:

Z = (observed percentage - expected percentage)

Standard Error of a percentage

Some differences: you are now working with a Standard Error of a percentage instead of a standard error of a sum. And the expected sum or percentage is a parameter and it is considered fixed and unchanging. The observed sum or percentage is a sample statistic and can change from sample to sample.

Example: for the sample of 400 students, we can say that the chance of getting between 140 (35%) and 180 (45%) students over age 25 is 95% and so forth.

Example: for the sample of 400 students, what is the chance of getting between 137 and 143 students over age 25?
Z = (143 - 160) / 10 = -17/10 = -1.7
Z =(137 - 160) / 10 = -23/10 = -2.3

A Z = -1.70 has % between +/- 1.7Z and a Z=-2.30 has % between +/- 2.3Z so chance of getting between 137 and 143 students is ((97.86-91.09)/2) = 3.385%

6. Things to note

(1) the parameter(s) stay fixed. For example 3, the box always has an average of .40 and the Standard Deviation of the box is always .50.
(2) the standard error for the sum (or count) of students gets larger as the sample gets larger
(3) as a percentage of the sample, the SE gets smaller, i.e. 5/100 = 5%, 15/900 = 1.7%, 30 / 3600 = 0.8%

Sample Size	Expected number	SE of the number	Expected percentage	SE of the percentage
100	40	5	40%	5%
400	160	10	40%	2.5%
900	360	15	40%	1.7%
3600	1440	30	40%	0.8%